



# A permutation model for free random variables and its classical analogue

Florent Benaych-Georges, Ion Nechita

## ► To cite this version:

Florent Benaych-Georges, Ion Nechita. A permutation model for free random variables and its classical analogue. 2009. hal-00220460v3

**HAL Id: hal-00220460**

**<https://hal.science/hal-00220460v3>**

Preprint submitted on 11 Feb 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A PERMUTATION MODEL FOR FREE RANDOM VARIABLES AND ITS CLASSICAL ANALOGUE

FLORENT BENAYCH-GEORGES AND ION NECHITA

**ABSTRACT.** In this paper, we generalize a permutation model for free random variables which was first proposed by Biane in [B95b]. We also construct its classical probability analogue, by replacing the group of permutations with the group of subsets of a finite set endowed with the symmetric difference operation. These constructions provide new discrete approximations of the respective free and classical Wiener chaos. As a consequence, we obtain explicit examples of non random matrices which are asymptotically free or independent. The moments and the free (resp. classical) cumulants of the limiting distributions are expressed in terms of a special subset of (noncrossing) pairings. At the end of the paper we present some combinatorial applications of our results.

## INTRODUCTION

Free probability is the non-commutative probability theory built upon the notion of independence called freeness. In classical probability theory, independence characterizes families of random variables whose joint distribution can be deduced from the individual ones by making their tensor product. In the same way, freeness, in free probability theory, characterizes families of random variables whose joint distribution can be deduced from the individual ones by making their free product (with the difference that free random variables belong to non commutative probability spaces, and that their joint distribution is no longer a probability measure, but a linear functional on a space of polynomials). Concretely, independent random variables are numbers arising randomly with no influence on each other, whereas free random variables are elements of an operator algebra endowed with a state which do not satisfy any algebraical relation together, as far as what can be observed with the algebra's state is concerned. Free probability theory has been a very active field of mathematics during the last two decades, constructed in a deep analogy with classical probability theory. It follows that there is a kind of dictionary between objects of both theories: many fundamental notions or results of classical probability theory, like the Law of Large Numbers, Central Limit Theorem, Gaussian distribution, convolution, cumulants, infinite divisibility have a precise analogue in free probability theory. Moreover, several examples of asymptotically free random variables have been found, like random matrices [V91, VDN92, HP00, HT05], representations of groups [B95a, B98], and a permutation model of P. Biane [B95b]. In the present paper, we shall firstly generalize this permutation model, and then develop its analogue from classical probability theory, which will allow us to show that surprisingly, in the "dictionary" mentioned above between classical and free probability theories, there is a correspondence (of minor importance when compared to others, but still interesting) between groups of sets endowed with the symmetric difference operation and groups of permutations, following from the correspondence between the lattice of partitions and the lattice of non crossing partitions.

---

2000 *Mathematics Subject Classification.* Primary 46L54, 60F05, Secondary 05A05, 05A18, 60H05 .

*Key words and phrases.* free probability, limit theorems, combinatorics of partitions, stochastic measures.

To explain how we construct this model and its analogue from classical probability theory, let us recall a few basic definitions of non commutative probability theory. First of all, let us recall that a *non commutative probability space* (as we shall use it) is a complex unital  $*$ -algebra  $\mathcal{A}$  endowed with a linear form  $\varphi$  such that  $\varphi(1) = 1$  and for all  $x \in \mathcal{A}$ ,  $\varphi(x^*) = \overline{\varphi(x)}$  and  $\varphi(xx^*) \geq 0$ . The *non commutative distribution* of a family  $(x_i)_{i \in I}$  of self-adjoint elements of  $\mathcal{A}$  is then the application which maps any polynomial  $P$  in the non commutative variables  $(X_i)_{i \in I}$  to  $\varphi(P((x_i)_{i \in I}))$ . This formalism is the one of free probability theory, but it recovers the one of classical probability theory, because if the algebra  $\mathcal{A}$  is commutative, then this distribution is actually the integration with respect to a probability measure on  $\mathbf{R}^I$ , and  $\mathcal{A}$  and  $\varphi$  can respectively be identified with a subalgebra of the intersection of the  $L^p$  spaces ( $p \in [1, +\infty)$ ) of a certain probability space and with the integration with respect to the probability measure of this probability space. A general example of non commutative probability space of historical importance is, given a countable group  $G$ , the  $*$ -algebra  $\mathbf{C}[G] = \mathbf{C}^{(G)}$  (the set of finitely supported complex functions on  $G$ ), endowed with the notion of adjoint defined by  $(\sum_{g \in G} x_g \cdot g)^* = \sum_{g \in G} \overline{x_g} \cdot g^{-1}$  and with the trace  $\varphi(\sum_{g \in G} x_g \cdot g) = x_e$ , where  $e$  denotes the neutral element of  $G$ . Our asymptotic model for free random variables is constructed in the algebra of the group  $\mathcal{S}$  of permutations with finite support of the set of nonnegative integers, whereas its classical probability theory analogue is constructed in the algebra of the group  $\mathcal{G}$  of finite sets of nonnegative integers endowed with the symmetric difference operation. More precisely, let us define, for all integer  $r \geq 1$ , and  $t \in [0, +\infty)$ , the element of  $\mathbf{C}[\mathcal{S}]$

$$M_r(n, t) = \frac{1}{n^{r/2}} \sum \underbrace{(0a_1a_2 \cdots a_r)}_{\substack{\text{designs the cycle} \\ 0 \rightarrow a_1 \rightarrow a_2 \rightarrow \cdots \rightarrow a_r \rightarrow 0}},$$

where the sum runs over all  $r$ -uplets  $(a_1, \dots, a_r)$  of pairwise distinct integers of  $[1, nt]$ . In [B95b], it was already proved that the non commutative distribution of the family  $(M_1(n, t))_{t \in [0, +\infty)}$  converges, as  $n$  goes to infinity, to the one of a family  $(M_1(t))_{t \in [0, +\infty)}$  which is a free Brownian motion. Here, we shall prove that the non commutative distribution of the family  $(M_r(n, t))_{r \geq 1, t \in [0, +\infty)}$  converges, as  $n$  goes to infinity, to the one of a family  $(M_r(t))_{r \geq 1, t \in [0, +\infty)}$  such that that for all  $r, t$ , one has  $M_r(t) = t^{\frac{r}{2}} U_r(t^{-1/2} M_1(t))$ , where the  $U_r$ 's are the Chebyshev polynomials of second kind.

The classical probability analogue of this model is constructed replacing the group  $\mathcal{S}$  of finitely-supported permutations of the set of nonnegative integers by the group  $\mathcal{G}$  of finite sets of nonnegative integers, endowed with the symmetric difference operation (the *symmetric difference*  $A \Delta B$  of two sets  $A$  and  $B$  is  $(A \cup B) \setminus (A \cap B)$ ), we define, for all integer  $r \geq 1$ , and  $t \in [0, +\infty)$ , the element of  $\mathbf{C}[\mathcal{G}]$

$$L_r(n, t) = \frac{1}{n^{r/2}} \sum \{a_1, a_2, \dots, a_r\},$$

where the sum runs over all  $r$ -uplets  $(a_1, \dots, a_r)$  of pairwise distinct integers of  $[1, nt]$ . We shall prove that the non commutative distribution of the family  $(L_r(n, t))_{r \geq 1, t \in [0, +\infty)}$  converges, as  $n$  goes to infinity, to the one of a family  $(L_r(t))_{r \geq 1, t \in [0, +\infty)}$  such that  $(L_1(t))_{t \in [0, +\infty)}$  is a classical Brownian motion and that for all  $r, t$ , one has  $L_r(t) = t^{\frac{r}{2}} H_r(t^{-1/2} L_1(t))$ , where the  $H_r$ 's are the Hermite polynomials.

This model is constructed on a commutative algebra, hence the joint distribution of the family  $(L_r(n, t))_{r \geq 1, t \in [0, +\infty)}$  is the one of a family of classical random variables, and the same holds for the family  $(L_r(t))_{r \geq 1, t \in [0, +\infty)}$ . This last process is well known in classical probability theory: up to a renormalization, it is the *Wiener chaos* [J97, N06, N09]. Our model provides a new discrete approximation of the random process  $(L_r(t))_{r \geq 1, t \in [0, +\infty)}$ . The Wiener chaos is the starting point of a wide construction of stochastic processes, called the *stochastic measures* [RW97, N06, PT08].

In a forthcoming paper, we shall prove that both our classical and free models can be generalized to this setting.

Let us now go back to the free model and explain how one can obtain non-random asymptotically free matrices. From the results stated earlier, there emerges the general idea that duly renormalized elements of  $\mathbf{C}[\mathcal{S}]$  of the type

$$A(n) := \sum_{\substack{a_1, \dots, a_{r_a} \\ \text{in a set of size } n}} (0a_1 \cdots a_{r_a}), \quad B(n) := \sum_{\substack{b_1, \dots, b_{r_b} \\ \text{in a set of size } n}} (0b_1 \cdots b_{r_b}), \quad C(n) := \sum_{\substack{c_1, \dots, c_{r_c} \\ \text{in a set of size } n}} (0c_1 \cdots c_{r_c}), \text{ etc.}$$

are asymptotically free as  $n$  goes to infinity if the respective sets where the  $a_i$ 's, the  $b_i$ 's and the  $c_i$ 's are picked from are pairwise disjoint, and that in this result, asymptotic freeness is replaced by asymptotic independence if the group  $\mathcal{S}$  of permutations is replaced by the one of finite sets endowed with the symmetric difference operation and cycles  $(0x_1 \cdots x_r)$  are replaced by sets  $\{x_1, \dots, x_r\}$ .

Let us now comment on Biane's original motivation for this construction. His idea (for  $r = 1$ ) easily generalizes for arbitrary  $r$ . As before, consider a finite set of elements  $A(n), B(n), C(n)$ , etc. of the group algebra  $\mathbf{C}[\mathcal{S}_N]$ , which is possible for  $N$  large enough. When viewed as operators on  $\mathcal{S}_N$ ,  $A(n), B(n), C(n)$ , etc. are complex matrices with rows and columns indexed by the elements of  $\mathcal{S}_N$  (these matrices can be seen as the adjacency matrices of some Cayley graphs). This is the reason why these results provide explicit examples of asymptotically free families of non random matrices. To our knowledge, there are no other such constructions. The classical probability part of our result also provides an explicit example of commutative family of non random matrices which are asymptotically independent, property that only random matrices had until now been proved to have.

In the last part of this paper, we shall explore connections between several combinatorial structures and the sets of non crossing pairings which appeared in the formulas of moments and free cumulants in the limit theorems presented above.

## 1. THE PERMUTATION MODEL FOR FREE RANDOM VARIABLES

**1.1. Computation of the limit distribution.** The non-commutative probability space we are going to work with is the group algebra  $\mathbf{C}[\mathcal{S}]$  of the group  $\mathcal{S}$  of finitely-supported permutations of the set of nonnegative integers (i.e. permutations for which all but finitely-many points are fixed points), with its canonical trace defined by  $\varphi(\sum_{\sigma} x_{\sigma} \sigma) = x_{id}$ , where  $id$  is the identity permutation. Let us define, for all integer  $r \geq 1$ , and  $t \in [0, +\infty)$ , the element of  $\mathbf{C}[\mathcal{S}]$

$$M_r(n, t) = \frac{1}{n^{r/2}} \sum (0a_1 a_2 \cdots a_r),$$

where the sum runs over all  $r$ -uplets  $(a_1, \dots, a_r)$  of pairwise distinct integers of  $[1, nt]$ . For  $r = 0$ , we put  $M_0(n, t) = id$ . Our purpose in what follows is to study the asymptotic properties (in the limit  $n \rightarrow \infty$ ) of the family  $(M_r(n, t))_{r,t}$ .

Before stating the main result of this section, let us recall that a free Brownian motion is a process  $(S_t)_{t \in [0, +\infty)}$  of non commutative random variables with free increments such that for all  $t$ ,  $S_t$  is semi-circular with variance  $t$ . Let us also recall some facts about the Chebyshev polynomials of the second kind, denoted by  $(U_n)$ . These are the orthogonal polynomials on  $[-2, 2]$  with respect to the semi-circular weight  $w(x) = \frac{1}{2\pi} \sqrt{4 - x^2}$ . They satisfy the property

$$U_n(2 \cos \theta) = \frac{\sin(n+1)\theta}{\sin \theta}, \quad \forall n \geq 0$$

and the recurrence relation

$$U_0(x) = 1, \quad U_1(x) = x, \quad U_1(x)U_n(x) = U_{n-1}(x) + U_{n+1}(x), \quad \forall n \geq 1.$$

**Theorem 1.** *The non commutative distribution of the family  $(M_r(n, t))_{r \geq 1, t \in [0, +\infty)}$  converges, as  $n$  goes to infinity, to the one of a family  $(M_r(t))_{r \geq 1, t \in [0, +\infty)}$  such that  $(M_1(t))_{t \in [0, +\infty)}$  is a free Brownian motion and for all  $r, t$ , one has  $M_r(t) = t^{\frac{r}{2}} U_r(t^{-1/2} M_1(t))$ , where the  $U_r$ 's are the Chebyshev polynomials of second kind.*

*Proof. Step I.* It follows from a direct application of Theorem 1 of [B95b] that the non commutative distribution of the family  $(M_1(n, t))_{t \in [0, +\infty)}$  converges, as  $n$  goes to infinity, to the one of a family  $(M_1(t))_{t \in [0, +\infty)}$  which is a free Brownian motion.

*Step II.* Let us prove that for all integer  $r \geq 1$ , and  $t \in (0, +\infty)$ ,

$$(1) \quad \lim_{n \rightarrow \infty} \varphi[(M_1(n, t)M_r(n, t) - tM_{r-1}(n, t) - M_{r+1}(n, t))^2] = 0.$$

We first compute  $M_1(n, t)M_r(n, t)$ :

$$\begin{aligned} M_1(n, t)M_r(n, t) &= n^{-\frac{r+1}{2}} \sum_{\substack{(a_1, \dots, a_r) \\ (a_{r+1})}} (0a_{r+1})(0a_1a_2 \cdots a_r) \\ &= n^{-\frac{r+1}{2}} \sum_{(a_1, \dots, a_{r+1})} (0a_1a_2 \cdots a_ra_{r+1}) + n^{-\frac{r+1}{2}} \sum_{k=1}^r \sum_{(a_1, \dots, a_r)} (0a_1a_2 \cdots a_{k-1})(a_k \cdots a_r) \\ &= M_{r+1}(n, t) + \frac{\lfloor nt \rfloor}{n} M_{r-1}(n, t) + n^{-\frac{r+1}{2}} \sum_{k=1}^{r-1} \sum_{(a_1, \dots, a_r)} (0a_1a_2 \cdots a_{k-1})(a_k \cdots a_r). \end{aligned}$$

Thus, it suffices to show that  $(a = (a_1, \dots, a_r), b = (b_1, \dots, b_r))$

$$\lim_{n \rightarrow \infty} \varphi[(n^{-\frac{r+1}{2}} \sum_{k=1}^{r-1} \sum_a (0a_1a_2 \cdots a_{k-1})(a_k \cdots a_r))^2] = 0.$$

But

$$\left( \sum_{k=1}^{r-1} \sum_a (0a_1a_2 \cdots a_{k-1})(a_k \cdots a_r) \right)^2 = \sum_{k,l=1}^{r-1} \sum_{a,b} (0a_1a_2 \cdots a_{k-1})(a_k \cdots a_r)(0b_1b_2 \cdots b_{l-1})(b_l \cdots b_r)$$

In order for the permutation on the right-hand side to be the identity, it has to be that

$$(0b_1b_2 \cdots b_{l-1})(b_l \cdots b_r) = [(0a_1a_2 \cdots a_{k-1})(a_k \cdots a_r)]^{-1} = (a_k a_r a_{r-1} \cdots a_{k+1})(0a_{k-1} \cdots a_1)$$

and thus  $k = l$  and the  $b$ 's are determined (modulo some circular permutation of size at most  $r$ ) by the  $a$ 's. We find that there are at most  $(r-1)r!(nt)^r$  terms in the sum which are equal to the identity and (1) follows.

*Step III.* To prove the existence of a limit to the non commutative distribution of the family  $(M_r(n, t))_{r \geq 1, t \in [0, +\infty)}$ , we have to prove that for all polynomial  $P$  in the non commutative variables  $(X_r(t))_{r \geq 0, t \in [0, +\infty)}$ ,

$$\varphi(P((M_r(n, t))_{r \geq 0, t \in [0, +\infty)}))$$

has a finite limit as  $n$  goes to infinity. First of all, by linearity, can suppose that  $P$  is a monomial  $X_{r_1}(t_1) \cdots X_{r_k}(t_k)$  with  $r_1, \dots, r_k \geq 0, t_1, \dots, t_k \in [0, +\infty)$ . Then let us prove it by induction on  $R := \max\{r_1, \dots, r_k\}$ . If  $R = 0$  or 1, it follows from the first step of the proof and the convention  $M_0(n, t) = 1$ . Now, let us suppose the result to be proved until rank  $R - 1$ . Replacing, for all  $t \in [0, +\infty)$ , each  $X_R(t)$  in  $P$  by

$$(X_1(t)X_{R-1}(t) - tX_{R-2}(t)) - (X_1(t)X_{R-1}(t) - tX_{R-2}(t) - X_R(t))$$

and using the second step of the proof with the Cauchy-Schwarz inequality, one gets the convergence. Let us denote the limit distribution by  $\Psi : \mathbf{C}\langle X_r(t) ; r \geq 0, t \in [0, +\infty) \rangle \rightarrow \mathbf{C}$ .

*Step IV.* Now, it remains only to identify the limit distribution. Note first that by the first step and the convention  $M_0(n, t) = 1$ , the Cauchy-Schwarz inequality allows us to claim that the bilateral ideal generated by

$$\{X_0(t) - 1 ; t \in [0, +\infty)\} \cup \{X_1(t)X_{m-1}(t) - tX_{m-2}(t) - X_m(t) ; m \geq 2, t \in [0, +\infty)\}$$

is contained in the kernel of  $\Psi$ . Hence up to a quotient of the algebra  $\mathbf{C}\langle X_r(t) ; r \geq 0, t \in [0, +\infty) \rangle$ , one can suppose that for all  $m \geq 2, t \in [0, +\infty)$ ,  $X_0(t) = 1$  and  $X_1(t)X_{m-1}(t) = tX_{m-2}(t) + X_m(t)$ . It allows us to claim that for all  $m \geq 0, t \in [0, +\infty)$ ,  $X_m(t)$  is a polynomial in  $X_1(t)$ , namely that  $X_m(t) = t^{\frac{m}{2}} U_m(t^{-1/2} X_1(t))$ , where the  $U_m$ 's are the Chebyshev polynomials of second kind (indeed, this family is completely determined by the fact that  $U_0 = 1, U_1 = X$  and for all  $m \geq 2, U_1 U_{m-1} = U_{m-2} + U_m$ ). Since by the first step,  $(M_1(t))_{t \in [0, +\infty)}$  is a free Brownian motion, the proof is complete.  $\square$

The following corollary generalizes Theorem 1 of [B95b]. Roughly speaking, it states that duly renormalized elements of  $\mathbf{C}[\mathcal{S}]$  of the type

$$A(n) := \sum_{\substack{a_1, \dots, a_{r_a} \\ \text{in a set of size } n}} (0a_1 \cdots a_{r_a}), \quad B(n) := \sum_{\substack{b_1, \dots, b_{r_b} \\ \text{in a set of size } n}} (0b_1 \cdots b_{r_b}), \quad C(n) := \sum_{\substack{c_1, \dots, c_{r_c} \\ \text{in a set of size } n}} (0c_1 \cdots c_{r_c}), \text{ etc.}$$

are asymptotically free as  $n$  goes to infinity if the respective sets where the  $a_i$ 's, the  $b_i$ 's and the  $c_i$ 's are picked are pairwise disjoint. Biane had proved it in the case where  $r_a = r_b = r_c = \dots = 1$ .

**Corollary 1.** *Fix  $p \geq 1, r_1, \dots, r_p \geq 0, t_0 < t_1 < \dots < t_p$ , and define, for all  $i = 1, \dots, p$ , for each  $n \geq 1, M_i(n) = n^{-\frac{r_i}{2}} \sum (0a_1 \cdots a_{r_i})$ , where the sum runs over all  $r_i$ -uplets  $(a_1, \dots, a_{r_i})$  of pairwise distinct integers of  $(nt_{i-1}, nt_i]$ . Then  $M_1(n), \dots, M_p(n)$  are asymptotically free as  $n$  goes to infinity.*

*Proof.* Let us define, for all  $i = 1, \dots, p$  and  $n \geq 1, S_i(n) := n^{-\frac{1}{2}} \sum_{\substack{a \in (nt_{i-1}, nt_i] \\ a \text{ integer}}} (0a)$ . By the

previous theorem, as  $n$  goes to infinity, the non commutative distribution of  $(S_1(n), \dots, S_p(n))$  tends to the one of a free family  $(s_1, \dots, s_p)$  of semi-circular elements (with various variances). Moreover, the same theorem says that for all  $i$ , as  $n$  goes to infinity,

$$\lim_{n \rightarrow \infty} \varphi((M_i(n) - (t_i - t_{i-1})^{\frac{r_i}{2}} U_{r_i}(\sqrt{t_i - t_{i-1}} S_i(n))^2) = 0.$$

It follows that the non commutative distribution of the family

$$(S_1(n), M_1(n), \dots, S_p(n), M_p(n))$$

converges to the one of

$$(s_1, (t_1 - t_0)^{\frac{r_1}{2}} U_{r_1}(\sqrt{t_1 - t_0} s_1), \dots, s_p, (t_p - t_{p-1})^{\frac{r_p}{2}} U_{r_p}(\sqrt{t_p - t_{p-1}} s_p),$$

which finishes the proof.  $\square$

**1.2. Moments and cumulants of the limit distribution.** We now turn to the moments and the free cumulants of the family  $(M_r(t))_{r \geq 1, t \in [0, +\infty)}$ . As we shall see, these quantities have elegant closed expressions in terms of non-crossing pairings of a special kind. Let us now introduce the combinatorial objects of interest. For  $f$  function defined on a finite set  $X$ ,  $\ker f$  designates the partition of  $X$  by the level sets of  $f$ . For every  $p \geq 1$  and for every vector  $r = (r_1, \dots, r_p)$  of positive integers, consider the function  $f_r : \{1, \dots, |r|\} \rightarrow \{1, \dots, p\}$  defined by  $f_r(x) = k$  if and only if  $r_1 + \dots + r_{k-1} < x \leq r_1 + \dots + r_k$  (we put  $|r| = r_1 + \dots + r_p$ ). We introduce the set

$NC_2(r)$  of non-crossing pairings  $\pi$  of the set  $\{1, \dots, |r|\}$  which do not link two elements who have the same image by  $f_r$ , i.e. such that  $\pi \wedge \hat{1}_r = \hat{0}_{|r|}$ , where  $\hat{1}_r = \ker f_r$  and  $\hat{0}_{|r|}$  is the singletons partition of  $\{1, \dots, |r|\}$ . We also introduce  $NC_2^*(r) = \{\pi \in NC_2(r) | \pi \vee \hat{1}_r = \hat{1}_{|r|}\}$ , where  $\hat{1}_{|r|}$  is the one-block-partition of  $\{1, \dots, |r|\}$ . For  $s$  positive integer, we denote with  $\langle s \rangle_p = (s, s, \dots, s)$  the constant vector where  $s$  appears  $p$  times.

In the following theorem, we compute the mixed moments and free cumulants of the family  $(M_r)_{r \geq 1} = (M_r(1))_{r \geq 1}$  (the mixed moments and cumulants of the family  $(M_r(t))_{r \geq 1, t \in [0, +\infty)}$  can easily be computed in the same way).

**Theorem 2.** *The distribution of the family  $(M_r)_{r \geq 1}$  is characterized by the fact that its mixed moments are given by*

$$\varphi(M_{r_1} M_{r_2} \cdots M_{r_p}) = \sharp NC_2(r)$$

*and its free cumulants are given by*

$$\kappa_p(M_{r_1}, M_{r_2}, \dots, M_{r_p}) = \sharp NC_2^*(r).$$

**Remark 1.** *Although they are clearly dependent, the elements  $M_r$  are not correlated:  $\varphi(M_q M_r) = 0$  if  $q \neq r$  (this follows from the orthogonality of the Chebyshev polynomials).*

**Remark 2.** *This theorem provides a new proof (even though there are already many !) of the formula of the free cumulants of the free Poisson distribution (also called Marchenko-Pastur distribution, see [HP00]). Indeed,  $M_2 + 1 = M_1^2$  is well known to have a free Poisson distribution with mean 1, all of whose cumulants except the first one the same as the free cumulants of  $M_2$ . By the theorem, for all  $p \geq 2$ ,  $\kappa_p(M_2)$  is the cardinality of  $\{\pi \in NC_2(2p) | \pi \vee \hat{1}_{\langle 2 \rangle_p} = \hat{1}_{2p}\}$ . In [NS06], it is shown that*

$$\{\pi \in NC(2p) | \pi \vee \hat{1}_{\langle 2 \rangle_p} = \hat{1}_{2p}\} = \{\pi \in NC(2p) | 1 \stackrel{\pi}{\sim} 2p, 2i \stackrel{\pi}{\sim} 2i+1, \forall i \in \{1, \dots, p-1\}\}.$$

Thus,

$$\{\pi \in NC_2(2p) | \pi \vee \hat{1}_{\langle 2 \rangle_p} = \hat{1}_{2p}\} = \{\{ \{2p, 1\}, \{2, 3\}, \dots, \{2p-2, 2p-1\} \}\},$$

which is a partition of  $NC_2(\langle 2 \rangle_p)$ , hence  $\kappa_p(M_2) = 1$ .

*Proof.* Let us first prove that the mixed moments are given by the formula of the theorem. Using the identity  $(0b_1 b_2 \cdots b_s) = (0b_s)(0b_{s-1}) \cdots (0b_1)$ , we have

$$\prod_{j=1}^p M_{r_j}(n, 1) = n^{-\frac{|r|}{2}} \sum_a (0a_1)(0a_2) \cdots (0a_{|r|}),$$

where the sum is taken over all families  $a = (a_1, \dots, a_{|r|}) \in \{1, \dots, n\}^{|r|}$  such that for all  $k, l \in \{1, \dots, |r|\}$ ,  $a_k \neq a_l$  whenever  $f_r(k) = f_r(l)$ . To such a family  $a$  we associate the partition  $\mathcal{P}(a)$  of the set  $\{1, \dots, |r|\}$  defined by  $k \sim l$  if and only if  $a_k = a_l$ . Thus, for all  $a$ ,  $\mathcal{P}(a)$  does not link two elements that have the same image by  $f_r$ , i.e. satisfies  $\mathcal{P}(a) \wedge \hat{1}_r = \hat{0}_{|r|}$ . We regroup the terms of the preceding sum according to the partitions  $\mathcal{P}$ :

$$\sum_{\pi} n^{-\frac{|r|}{2}} \sum_{a: \mathcal{P}(a) = \pi} (0a_1)(0a_2) \cdots (0a_{|r|}).$$

Let us show that among the partitions  $\pi$  such that  $\pi \wedge \hat{1}_r = \hat{0}_{|r|}$ , the only partitions that contribute to the limit, as  $n$  goes to infinity, are non-crossing pairings, i.e. elements of  $NC_2(r)$ . If  $\pi = \mathcal{P}(a)$  contains a singleton  $\{k\}$ , then the permutation  $(0a_1)(0a_2) \cdots (0a_{|r|})$  cannot be the identity, because the element  $a_k$  appears only once and thus its image cannot be itself. Consider now a partition  $\pi$  with no singleton but with a class with at least three elements. It is easy to show that there are no more than  $n^{\frac{|r|-1}{2}}$  families  $a$  such that  $\mathcal{P}(a) = \pi$  and thus they have

no contribution asymptotically. We have shown that only pairings contribute to the trace. The argument in [B95b], Lemma 2 (which adapts *mutatis mutandis* to our case) shows that only the non-crossing pairings contribute, completing the proof.

Let us now compute the free cumulants. To a pairing  $\mathcal{P} \in NC_2(r)$  we associate the non-crossing partition  $\bar{\mathcal{P}} \in NC(p)$  which encodes the way  $\mathcal{P}$  links the blocks of  $\hat{1}_r : k \stackrel{\bar{\mathcal{P}}}{\sim} l$  if and only if  $r_1 + \dots + r_k \stackrel{\mathcal{P}}{\sim} \hat{1}_r r_1 + \dots + r_l$ , for all  $k, l \in \{1, \dots, p\}$ . We have

$$\varphi((M_{r_1} M_{r_2} \cdots M_{r_p}) = \#NC_2(r) = \sum_{\pi \in NC(p)} \#\{\mathcal{P} \in NC_2(r) | \bar{\mathcal{P}} = \pi\}.$$

Since the functionals  $NC(p) \ni \pi \mapsto \#\{\mathcal{P} \in NC_2(r) | \bar{\mathcal{P}} = \pi\}$  are multiplicative, we have identified the free cumulants of the family  $(M_r)_{r \geq 1}$ :

$$\forall p \geq 1, r_1, \dots, r_p \geq 1, \kappa_\pi(M_{r_1}, M_{r_2}, \dots, M_{r_p}) = \#\{\mathcal{P} \in NC_2(r) | \bar{\mathcal{P}} = \pi\}.$$

Considering the case  $\pi = \hat{1}_p$ , we obtain the announced formula for the free cumulants.  $\square$

**1.3. An application: linearization coefficients for orthogonal polynomials.** As a corollary of Theorems 1 and 2, we recover some formulas already obtained in [A05] using different techniques. Consider a family  $(P_n)$  of orthonormal polynomials with respect to some weight  $w$ . For an integer vector  $r = (r_1, \dots, r_p)$  there is a decomposition

$$P_{r_1}(x) P_{r_2}(x) \cdots P_{r_p}(x) = \sum_{k=0}^{|r|} c_k^{(r)} P_k(x),$$

where the scalars  $c_k^{(r)} \in \mathbf{R}$  are called *linearization coefficients* of the family  $(P_n)$ . They can easily be recovered by integration:

$$c_k^{(r)} = \int P_{r_1}(x) P_{r_2}(x) \cdots P_{r_p}(x) \cdot P_k(x) dw(x).$$

For the Chebyshev polynomials, these integrals are the expectation (the trace) of the corresponding products of the random variables  $M_r$ :

**Corollary 2.** *The linearization coefficients for the Chebyshev polynomials of the second kind  $U_n$  are given by*

$$(2) \quad c_k^{(r)} = \#NC_2(r \cup k),$$

where  $r \cup k$  is the vector  $(r_1, \dots, r_p, k)$ .

**Remark 3.** *It was mentioned to us by a referee that formula (2) had already been proved, with another method, in [SCV84, Th. 7].*

In [A05], a similar formula is deduced for the centered free Charlier polynomials  $V_n$ . These polynomials are orthogonal with respect to the centered Marchenko-Pastur density

$$w_2(t) = \mathbf{1}_{[-1,3]}(t) \frac{1}{2\pi} \sqrt{\frac{4}{1+t} - 1}.$$

Note that  $M_2 = M_1^2 - 1$  has the distribution  $d\mu_2 = w_2(t)dt$ . Moreover, one can easily see that  $V_n \circ U_2 = U_{2n}$  and thus

$$\int V_{r_1}(x) V_{r_2}(x) \cdots V_{r_p}(x) \cdot V_k(x) dw_2(x) = \int U_{2r_1}(x) U_{2r_2}(x) \cdots U_{2r_p}(x) \cdot U_{2k}(x) dw(x).$$

We obtain



**Corollary 3.** *The linearization coefficients for the centered free Charlier polynomials  $V_n$  are given by*

$$d_k^{(r)} = \#NC_2(2r \cup 2k),$$

where  $2r \cup 2k$  is the vector  $(2r_1, \dots, 2r_p, 2k)$ .

Using the bijection between non-crossing pairings of size  $2n$  and non-crossing partitions of size  $n$  (see [NS06], pp. 153–154), one can easily see that the sets  $NC_2(2r \cup 2k)$  and  $\{\pi \in NC(r \cup k | \pi \text{ has no singleton})\}$  have the same cardinality, hence our formula is equivalent to the one in [A05].

## 2. A CLASSICAL PROBABILITY ANALOGUE

The model we have studied in the first part involves permutations, asymptotical freeness, non-crossing pairings, the semi-circular distribution and its orthogonal polynomials, the second kind Chebyshev polynomials. By replacing permutations with sets, we construct in this section an analogue model, where the objects from free probability are replaced by their classical counterparts, respectively independence, (possibly crossing) pairings, and the Gaussian distribution with the orthogonal Hermite polynomials.

**2.1. Computation of the limit distribution.** The non-commutative probability space we are going to work with here is the group algebra  $\mathbf{C}[\mathcal{G}]$  of the group  $\mathcal{G}$  of finite sets of nonnegative integers endowed with the symmetric difference operation, with its canonical trace defined by  $\psi(\sum_A x_A A) = x_\emptyset$ . Let us define, for all integer  $r \geq 1$ , and  $t \in [0, +\infty)$ , the element of  $\mathbf{C}[\mathcal{G}]$

$$L_r(n, t) = \frac{1}{n^{r/2}} \sum \{a_1, a_2, \dots, a_r\},$$

where the sum runs over all  $r$ -uplets  $(a_1, \dots, a_r)$  of pairwise distinct integers of  $[1, nt]$ . For  $r = 0$ , we put  $L_0(n, t) = \emptyset$  (which is the unity of this algebra). Our purpose in what follows is to study the asymptotic properties (in the limit  $n \rightarrow \infty$ ) of the family  $(L_r(n, t))_{r,t}$ .

Recall that for every  $p \geq 1$  and for every vector  $r = (r_1, \dots, r_p)$  of positive integers, the function  $f_r : \{1, \dots, |r|\} \rightarrow \{1, \dots, p\}$  is the projection defined by  $f_r(x) = k$  iff  $r_1 + \dots + r_{k-1} < x \leq r_1 + \dots + r_k$  ( $|r| = r_1 + \dots + r_p$ ). We replace the non-crossing partitions from the free case with general partitions:  $\Pi_2(r)$  is the set of pairings  $\pi$  of  $\{1, \dots, |r|\}$  which do not link two elements who have the same image by  $f_r$ , i.e. such that  $\pi \wedge \hat{1}_r = \hat{0}_{|r|}$ , where  $\hat{1}_r$  is still the partition of  $\{1, \dots, |r|\}$  with blocks  $f_r^{-1}(1), f_r^{-1}(2), \dots, f_r^{-1}(p)$ . We also introduce  $\Pi_2^*(r) = \{\pi \in \Pi_2(r) | \pi \vee \hat{1}_r = \hat{1}_{|r|}\}$ .

In the following lemma we compute the asymptotic joint moments of the random variables  $L_r(n, t)$ .

**Lemma 1.** *Let  $p \geq 1$  and consider  $t_1, \dots, t_p > 0$  and a family of positive integers  $r = (r_1, \dots, r_p)$ . Then, in the limit  $n \rightarrow \infty$ , the trace  $\psi \left[ \prod_{j=1}^p L_{r_j}(n, t_j) \right]$  converges to*

$$\sum_{\pi \in \Pi_2(r)} \prod_{\{i,j\} \in \pi} \min(t_{f_r(i)}, t_{f_r(j)}).$$

*Proof.* Using the properties of the symmetric difference  $\Delta$ , we get

$$\prod_{j=1}^p L_{r_j}(n, t_j) = n^{-\frac{|r|}{2}} \sum_a \{a_1\} \Delta \{a_2\} \Delta \dots \Delta \{a_{|r|}\},$$

where the sum is taken over all families  $a = (a_1, \dots, a_{|r|})$  of positive integers such that for all  $k, l \in \{1, \dots, |r|\}$ ,  $a_k \in [1, nt_{f_r(k)}]$  and  $a_k \neq a_l$  whenever  $f_r(k) = f_r(l)$ . To such a family  $a$  we associate the partition  $\mathcal{P}(a)$  of the set  $\{1, \dots, |r|\}$  defined by  $k \sim l$  if and only if  $a_k = a_l$ . Thus,

for all  $a$ ,  $\mathcal{P}(a)$  does not link two elements that have the same image by  $f_r$ . We regroup the terms of the preceding sum according to the partitions  $\mathcal{P}$ :

$$\sum_{\pi} n^{-\frac{|r|}{2}} \sum_{a: \mathcal{P}(a)=\pi} \{a_1\} \Delta \cdots \Delta \{a_{|r|}\}.$$

Let us show that only pairings can contribute to the asymptotic trace of the sum. It is obvious that  $\{a_1\} \Delta \cdots \Delta \{a_{|r|}\}$  is the empty set if and only if each  $a_i$  appears an even number of times. Thus,  $\pi = \mathcal{P}(a)$  cannot contain singletons. On the other hand, if  $\pi$  contains no singleton but has a class with at least three elements, it is easy to show that there are no more than  $(n \max\{t_1, \dots, t_p\})^{\frac{|r|-1}{2}}$  families  $a$  such that  $\mathcal{P}(a) = \pi$  and thus such partitions  $\pi$  do not contribute asymptotically.

For  $\pi$  pairing of  $\Pi_2(r)$ , the number of families  $a$  such that  $\mathcal{P}(a) = \pi$ , is asymptotic, as  $n$  goes to infinity, to  $n^{\frac{|r|}{2}} \prod_{\{i,j\} \in \pi} \min(t_{f_r(i)}, t_{f_r(j)})$ , which concludes the proof.  $\square$

Before stating the main result of this section, let us recall some facts about the Hermite polynomials, denoted by  $(H_n)$ . These are the orthogonal polynomials on the real line with respect to the standard Gaussian measure. They satisfy the recurrence relation

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_1(x)H_r(x) = H_{r+1}(x) + rH_{r-1}(x), \forall r \geq 1.$$

**Theorem 3.** *The distribution of the family  $(L_r(n, t))_{r \geq 1, t \in [0, +\infty)}$  converges, as  $n$  goes to infinity, to the one of a commutative family  $(L_r(t))_{r \geq 1, t \in [0, +\infty)}$  such that  $(L_1(t))_{t \in [0, +\infty)}$  is a classical Brownian motion and for all  $r, t$ , one has  $L_r(t) = t^{\frac{r}{2}} H_r(t^{-1/2} L_1(t))$ , where the  $H_r$ 's are the Hermite polynomials.*

*Proof. Step 0.* Note first that the symmetric difference is a commutative operation on sets. Hence the algebra  $\mathbf{C}[\mathcal{G}]$  is commutative.

*Step I.* It follows from a direct application of the previous lemma that the distribution of the family  $(L_1(n, t))_{t \in [0, +\infty)}$  converges, as  $n$  goes to infinity, to the one of a classical Brownian motion  $(L_1(t))_{t \in [0, +\infty)}$ .

*Step II.* Let us prove that for all integer  $r \geq 1$ , and  $t \in (0, +\infty)$ ,

$$(3) \quad \lim_{n \rightarrow \infty} \psi[(L_r(n, t)L_1(n, t) - rtL_{r-1}(n, t) - L_{r+1}(n, t))^2] = 0.$$

This is a consequence of the following computation of  $L_r(n, t)L_1(n, t)$ . The sums run over integers of  $[1, nt]$ .

$$\begin{aligned} L_r(n, t)L_1(n, t) &= n^{-\frac{r+1}{2}} \sum_{\substack{(a_1, \dots, a_r) \\ (a_{r+1})}} \{a_1\} \Delta \cdots \Delta \{a_{r+1}\} \\ &= n^{-\frac{r+1}{2}} \sum_{(a_1, \dots, a_{r+1})} \{a_1, a_2, \dots, a_r, a_{r+1}\} \\ &\quad + n^{-\frac{r+1}{2}} \sum_{k=1}^r \sum_{(a_1, \dots, a_r)} \{a_1, a_2, \dots, \check{a}_k, \dots, a_r\} \\ &= L_{r+1}(n, t) + n^{-\frac{r+1}{2}} \sum_{k=1}^r ([nt] - r + 1) \sum_{(b_1, \dots, b_{r-1})} \{b_1, b_2, \dots, b_{r-1}\} \\ &= L_{r+1}(n, t) + \frac{[nt] - r + 1}{n} r L_{r-1}(n, t) \\ &= L_{r+1}(n, t) + rt L_{r-1}(n, t) + \epsilon_n L_{r-1}(n, t), \text{ with } \epsilon_n \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

*Step III* and *Step IV* are as in the proof of Theorem 1, with the difference that here, the algebra is commutative, hence one-dimensional non-commutative distributions are integrations with respect to a probability measure, which is unique in this case.  $\square$

The following corollary is the classical probability theory counterpart of corollary 1. Roughly speaking, it states that duly renormalized elements of  $\mathbf{C}[\mathcal{G}]$  of the type

$$A(n) := \sum_{\substack{a_1, \dots, a_{r_a} \\ \text{in a set of size } n}} \{a_1, \dots, a_{r_a}\}, \quad B(n) := \sum_{\substack{b_1, \dots, b_{r_b} \\ \text{in a set of size } n}} \{b_1, \dots, b_{r_b}\}, \quad C(n) := \sum_{\substack{c_1, \dots, c_{r_c} \\ \text{in a set of size } n}} \{c_1, \dots, c_{r_c}\}, \dots$$

are asymptotically independent as  $n$  goes to infinity if the respective sets where the  $a_i$ 's, the  $b_i$ 's and the  $c_i$ 's are picked are pairwise disjoint.

**Corollary 4.** *Fix  $p \geq 1$ ,  $r_1, \dots, r_p \geq 0$ ,  $t_0 < t_1 < \dots < t_p$ , and defines, for all  $i = 1, \dots, p$ , for each  $n \geq 1$ ,  $L_i(n) = n^{-\frac{r_i}{2}} \sum \{a_1, \dots, a_{r_i}\}$ , where the sum runs over all  $r_i$ -uplets  $(a_1, \dots, a_{r_i})$  of pairwise distinct integers of  $(nt_{i-1}, nt_i]$ . Then  $L_1(n), \dots, L_p(n)$  are asymptotically independent as  $n$  goes to infinity.*

*Proof.* *Mutatis mutandis*, the proof goes along the same lines as the one of corollary 1.  $\square$

**2.2. Moments and cumulants of the limit distribution.** In the following theorem, we compute the mixed moments and cumulants of the family  $(L_r)_{r \geq 1} = (L_r(1))_{r \geq 1}$  (the mixed moments and cumulants of the family  $(L_r(t))_{r \geq 1, t \in [0, +\infty)}$  can easily be computed in the same way). Here, the analogy with the free probability model is obvious, since the formulas are the same ones as in Theorem 2, with the difference that the pairings are now allowed to have crossings.

**Theorem 4.** *The distribution of the family  $(L_r)_{r \geq 1}$  is characterized by the fact that its mixed moments are given by*

$$\psi(L_{r_1} L_{r_2} \cdots L_{r_p}) = \sharp \Pi_2(r)$$

*and its classical cumulants are given by*

$$k_p(L_{r_1}, L_{r_2}, \dots, L_{r_p}) = \sharp \Pi_2^*(r).$$

*Proof.* The moments have been computed in Lemma 1 and the cumulants can be computed in the same way as in the proof of Theorem 2.  $\square$

**Remark 4.** *The correspondence between the limit distributions of the classical and the free case is not the Bercovici-Pata bijection, since the distribution of  $L_2$  is not a classical Poisson distribution.*

**2.3. An application: linearization coefficients for orthogonal polynomials.** As in Corollaries 2 and 3, one deduce from this work combinatorial formulas for the linearization coefficients for Hermite and centered Charlier polynomials. Up to normalization, the formulas are the same ones, with the difference that non crossing pairings are replaced by pairings.

### 3. FURTHER COMBINATORICS

In this section, we explore connections between several combinatorial structures and the sets  $NC_2(r)$  and  $NC_2^*(r)$ , which appeared in the formulas of moments and free cumulants of the family  $M_r(1)$ .

**3.1. A bijection with a class of paths.** Here, we shall denote the set of nonnegative integers by  $\mathbf{N}$  and the set of integers by  $\mathbf{Z}$ .

It is well known that for all  $n \geq 1$ , the  $n$ -th moment of a semi-circular element is the number of Dyck paths with length  $n$ , i.e. of functions  $\gamma : \{0, \dots, n\} \rightarrow \mathbf{N}$  such that  $\gamma(0) = \gamma(n) = 0$  and for all  $i$ ,  $|\gamma(i) - \gamma(i-1)| = 1$ . Since for  $n, t$  fixed, the  $M_r(n, t)$ 's ( $r \geq 1$ ) are a generalizations of the Jucys-Murphy element  $M_1(n, t)$ , whose distribution tends to a semi-circular one, it is natural to expect a generalization of this interpretation of the moments in terms of paths for the moments of the  $M_r(t)$ 's. We show here that the mixed moments and free cumulants of the family  $(M_r)_{r \geq 1}$  count lattice paths with general jump size, as follows. Consider an integer vector  $r = (r_1, \dots, r_p)$ . For  $k \geq 1$ , define  $\Delta(k) = \{k, k-2, k-4, \dots, -k+2, -k\} = \{t-s; s, t \in \mathbf{N}, s+t=k\} \subset \mathbf{Z}$ . We define a *Dyck  $r$ -path* to be a function  $\gamma : \{0, 1, \dots, p\} \rightarrow \mathbf{Z}$  such that  $\gamma(0) = 0$ ,  $\gamma(p) = 0$ ,  $\gamma(i) + \gamma(i-1) \geq r_i$  and  $\gamma(i) - \gamma(i-1) \in \Delta(r_i)$  for all  $i \in \{1, \dots, p\}$  ( $\Delta(k)$  is somehow the set of admissible jumps for these paths). We denote by  $\Gamma(r)$  the set of Dyck  $r$ -paths and we also consider its subset  $\Gamma^*(r)$  of *irreducible* Dyck  $r$ -paths: a Dyck  $r$ -path  $\gamma$  is said to be irreducible if it has the property that it does not contain strictly smaller Dyck  $s$ -paths, in the following sense: there is no pair  $(x, y) \neq (0, p)$  such that the path  $\bar{\gamma} : \{0, \dots, y-x\} \rightarrow \mathbf{Z}$  defined by  $\bar{\gamma}(i) = \gamma(x+i) - \gamma(x)$  is a Dyck  $(r_{x+1}, r_{x+2}, \dots, r_y)$ -path.

It can be easily seen that Dyck  $r$ -paths are always positive ( $\gamma(i) \geq 0$ , for all  $i \in \{0, \dots, p\}$ ) and that the first and the last jumps are the largest, respectively smallest, possible:  $\gamma(1) = r_1$  and  $\gamma(p-1) = r_p$ . By the following proposition, Dyck  $r$ -paths (resp. irreducible ones) are counted by the moments (resp. free cumulants) of the family  $(M_r)_r := (M_r(1))_r$ :

**Proposition 1.** *The sets  $NC_2(r)$  and  $\Gamma(r)$  are in bijection. The same holds true for  $NC_2^*(r)$  and  $\Gamma^*(r)$ . In particular, we have*

$$\varphi(M_{r_1} M_{r_2} \cdots M_{r_p}) = \sharp \Gamma(r)$$

and

$$\kappa_p(M_{r_1}, M_{r_2}, \dots, M_{r_p}) = \sharp \Gamma^*(r).$$

*Proof.* Consider a non-crossing pairing  $\pi \in NC_2(r)$ . We begin by constructing the path of  $\Gamma(r)$  associated to  $\pi$ . An element  $k$  of  $\{1, \dots, |r|\}$  is said to be an *opener* (for  $\pi$ ) if it appears first in its block (pair) of  $\pi$ . Otherwise, it is called a *closer*. For  $1 \leq i \leq p$ , let  $B_i = f_r^{-1}(i)$ . As  $\pi$  is non-crossing and it does not contain pairs with both ends in  $B_i$ , the closers appear before the openers in each  $B_i$ . Let  $s_i$  be the number of closers of  $B_i$  and  $t_i$  be the number of openers of  $B_i$ . We have  $s_i + t_i = r_i$ . Define  $\gamma : \{0, 1, \dots, p\} \rightarrow \mathbf{Z}$  by  $\gamma(0) = \gamma(p) = 0$  and  $\gamma(i) - \gamma(i-1) = t_i - s_i$ , for all  $1 \leq i \leq p$ ; we have thus  $\gamma(i) - \gamma(i-1) \in \Delta(r_i)$ . The value of  $\gamma(i)$  is the number of *open* pairs after the first  $i$  groups of  $\pi$ . Hence, for all  $i \geq 1$ ,  $\gamma(i-1) - s_i \geq 0$ . This implies  $\gamma(i) + \gamma(i-1) \geq r_i$ , and thus  $\gamma$  is an  $r$ -path. In order to prove the other direction, note that a pairing  $\pi \in NC_2(r)$  can be reconstructed by knowing only the number of openers/closers in each block  $B_i$ . This information can easily be deduced from an  $r$ -path  $\gamma$ .

The proof that the construction above is a bijection between the set of irreducible  $r$ -paths  $\Gamma^*(r)$  and  $NC_2^*(r)$  is cumbersome; we shall just give the main idea. Again, let  $\pi$  be a pairing of  $NC_2(r)$ . The condition  $\pi \vee \hat{1}_r = \hat{1}_{|r|}$  amounts to the fact that the standard graphical representation of  $\pi$  and  $\hat{1}_r$  on the same figure ( $\hat{1}_r$  can drawn by connecting the points of each of its groups by horizontal lines) is a connected graph. If it is not the case, then the sub-graph of a connected component corresponds to a strictly smaller  $r$ -path in the path  $\gamma$  previously associated to  $\pi$ .  $\square$

**Remark 5.** *Note that for  $r = \langle 1 \rangle_p$ ,  $\Delta(1) = \{\pm 1\}$ , and we recover the usual Dyck paths. For  $r = \langle 2 \rangle_p$ , and  $p \geq 2$ , it can easily be seen that  $\Gamma^*(\langle 2 \rangle_p) = \{(0, 2, 2, \dots, 2, 0)\}$ , and we obtain the free cumulants of the centered Marchenko-Pastur (or free Poisson) distribution.*

**3.2. A Toeplitz algebra model for  $(M_r(1))_{r \geq 1}$ .** In this section we provide a concrete realization of the family  $(M_r(1))_{r \geq 1}$  by Toeplitz operators. Consider the Toeplitz algebra  $\mathcal{T}$  of bounded linear operators on  $\ell^2(\mathbf{N})$  with its vacuum state  $\omega(T) = \langle e_0, T e_0 \rangle$ . The shift operators are denoted by  $S$  and  $S^*$ . Let  $T_0 = 1$  and define, for all  $r \geq 1$  the operators

$$T_r = \sum_{k=0}^r \underbrace{SS \cdots S}_{r-k \text{ times}} \underbrace{S^* S^* \cdots S^*}_{k \text{ times}} = S^r + S^{r-1} S^* + \cdots + S^{*r}.$$

It can be easily checked that the operators  $T_r$  verify the recurrence relation of the (second kind) Chebyshev polynomials  $T_1 T_r = T_{r-1} + T_{r+1}$ . It is well known that, under the vacuum state, the operator  $T_1 = S + S^*$  has a semicircular distribution, and thus it has the same law as  $M_1(1)$ . We conclude that

**Proposition 2.** *The families  $(T_r)_r \in (\mathcal{T}, \omega)$  and  $(M_r(1))_r \in (\mathcal{A}, \varphi)$  have the same distribution.*

**Remark 6.** *Note that we can also realize the whole family  $(M_r(t))_{r \geq 1, t \in [0, +\infty)}$  on the full Fock space of the Hilbert space  $L^2([0, +\infty), dx)$  with the operators (here,  $\ell$  denotes the creation operator)*

$$T_r(t) = \sum_{k=0}^r \underbrace{\ell(\mathbf{1}_{[0,t]}) \cdots \ell(\mathbf{1}_{[0,t]})}_{r-k \text{ times}} \underbrace{\ell^*(\mathbf{1}_{[0,t]}) \cdots \ell^*(\mathbf{1}_{[0,t]})}_{k \text{ times}} \in \mathcal{B}(\mathcal{F}(L^2([0, +\infty), dx))).$$

It can be insightful to look at the matrix representations of the operators  $T_r$ . It can be easily verified that the  $(i, j)$  coefficient of  $T_r$ ,  $T_r(i, j) = \langle e_i, T_r e_j \rangle$  is null, unless

- $j - i \in \Delta(r) = \{r, r-2, \dots, -r\}$  and
- $j + i \geq r$ ,

in which case it equals 1.

This matrix point of view introduces the connection with the set  $\Gamma(r)$ :

$$\begin{aligned} \varphi(M_{r_1} M_{r_2} \cdots M_{r_p}) &= \omega(T_{r_1} T_{r_2} \cdots T_{r_p}) = [T_{r_1} T_{r_2} \cdots T_{r_p}](0, 0) = \\ &= \sum_{i_0=0, i_1, \dots, i_p=0} T_{r_1}(i_0, i_1) T_{r_2}(i_1, i_2) \cdots T_{r_p}(i_{p-1}, i_p). \end{aligned}$$

In order for the general term of the above sum to be non-zero, it has to be that each factor is 1, and that amounts to the fact that  $\gamma = (i_0, i_1, \dots, i_p) \in \Gamma(r)$ , providing a connection with the generalized Dyck paths discussed earlier.

**3.3. Non-commutative invariants and semi-standard Young tableaux.** In this section we show that the combinatorics of the family  $(M_r)_r$  is related to semi-standard Young tableaux, which have been shown to count the number of non-commutative classical invariants of binary forms [T88]. Here, we prove only a combinatorial result; whether there is a more profound reason for this, we ignore at this moment and connections with the representation theory of  $SL_2(\mathbf{C})$ ,  $GL(n)$  or  $S_n$  are to be explored.

Start by fixing a vector  $r = (r_1, \dots, r_p)$  such that  $|r|$  is even and consider the Young diagram with 2 rows and  $|r|/2$  columns associated to the partition  $\lambda = (|r|/2, |r|/2)$  of  $|r|$ . A semi-standard Young tableau of shape  $\lambda$  and weight  $r$  is a numbering of the Young diagram of shape  $\lambda$  with  $r_1$  1's,  $r_2$  2's,  $\dots$ ,  $r_p$   $p$ 's such that the rows are not decreasing and the columns are increasing. Let  $c(r)$  be the number of such semi-standard Young tableaux.

**Proposition 3.**  $c(r) = \sharp NC_2(r)$ .

*Proof.* We shall construct a bijection between the set of non-crossing pairings of  $NC_2(r)$  and the set of semi-standard Young tableaux of weight  $r$ . Start with a pairing  $\pi \in NC_2(r)$ . We shall add numbers in the empty Young diagram group by group. When we arrive at the  $i$ -th group of  $\pi$ , start by appending  $s_i$   $i$ 's to the second row, corresponding to the  $s_i$  closing pairs of the  $i$ -th group. Then add the remaining  $t_i$   $i$ 's to the top row - these are the  $t_i$  opening pairs. In this way we are sure to get a row non-decreasing numbering. The fact that the columns are increasing follows from the fact that at each moment, the number of opened pairs of  $\pi$  is larger or equal than the number of closed pairs. Thus the top row is always more occupied than the bottom row.  $\square$

**Remark 7.** As we did for the paths, we can prove a bijection between  $NC_2^*(r)$  and a strict subset of semi-standard Young tableaux. However, this is stricter than the notion of “indecomposable” Young tableaux, defined in [T88].

## REFERENCES

- [A05] M. Anshelevich, *Linearization coefficients for orthogonal polynomials using stochastic processes*, Ann. Probab., **33** (2005), no. 1, pp. 114–136.
- [B95a] Biane, Philippe *Representations of unitary groups and free convolution*, Publ. Res. Inst. Math. Sci. 31 (1995), no. 1, 63–79.
- [B95b] Biane, Philippe *Permutation model for semi-circular systems and quantum random walks*, Pacific J. Math. 171 (1995), no. 2, 373–387.
- [B98] Biane, Philippe *Representations of symmetric groups and free probability*, Adv. Math. 138 (1998), no. 1, 126–181.
- [HT05] Haagerup, Uffe; Thorbjørnsen, Steen *A new application of random matrices:  $\text{Ext}(C_{\text{red}}^*(F_2))$  is not a group*, Ann. of Math. (2) 162 (2005), no. 2, 711–775
- [HP00] Hiai, F., Petz, D. *The semicircle law, free random variables, and entropy*, Amer. Math. Soc., Mathematical Surveys and Monographs Volume **77**, 2000
- [J97] S. Janson *Gaussian Hilbert spaces*. Cambridge Tracts in Mathematics, 1997.
- [NS06] A. Nica, R. Speicher, *Lectures on the combinatorics of free probability*, Cambridge University Press, 2006.
- [N06] D. Nualart *The Malliavin calculus and related topics of Probability and Its Applications*. Springer Verlag, Berlin, Second edition, 2006.
- [N09] I. Nourdin *Integration by parts on Wiener space and applications*. Notes available on the web page of the author.
- [PT08] G. Peccati, M.S. Taqqu *Moments, cumulants and diagram formulae for non-linear functionals of random measures*. Survey, available on the web page of the first named author.
- [RW97] G.-C. Rota and C. Wallstrom *Stochastic integrals: a combinatorial approach*. The Annals of Probability 25(3), 1257–1283 (1997).
- [SCV84] M. de Sainte-Catherine, G. Viennot, *Combinatorial interpretation of products of Hermite, Laguerre and Tchebicheff polynomials* in Lecture Notes in Math. 1171, Polynomes Orthogonaux et Applications, Proceedings, Bar-le-Duc, 1984. 120–128.
- [T88] Y. Teranishi, *Noncommutative classical invariant theory*, Nagoya Math. J., 112 (1988), 153–169.
- [V91] Voiculescu, D.V. *Limit laws for random matrices and free products*, Invent. Math. **104** (1991) 201–220
- [VDN92] Voiculescu, D.V., Dykema, K., Nica, A. *Free random variables*, CRM Monographs Series No.1, Amer. Math. Soc., Providence, RI, 1992

LPMA, UPMC UNIV PARIS 6, CASE COURIER 188, 4, PLACE JUSSIEU, 75252 PARIS CEDEX 05, FRANCE.  
E-mail address: florent.benaych@gmail.com

UNIVERSITÉ DE LYON, UNIVERSITÉ DE LYON 1, INSTITUT CAMILLE JORDAN, CNRS UMR 5208, 43, BOULEVARD DU 11 NOVEMBRE 1918, 69622 VILLEURBANNE CEDEX, FRANCE  
E-mail address: nechita@math.univ-lyon1.fr